

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: August 22, 2003, 15:13:59 ; Search time 97 Seconds  
(without alignments)  
931.118 Million cell updates/sec

Title: US-09-745-506-37  
Perfect score: 350  
Sequence: 1 MDKALLSLINDFASLSFAE.....LENKINIILSETDRPFLQVY 350

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 830525 seqs, 258052604 residues

Word size : 0

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

SPTREMBL\_23:\*

- 1: sp-archaea:\*
- 2: sp-bacteria:\*
- 3: sp-fungi:\*
- 4: sp-human:\*
- 5: sp-invertebrate:\*
- 6: sp-mammal:\*
- 7: sp-mhc:\*
- 8: sp-organellar:\*
- 9: sp-phage:\*
- 10: sp-plant:\*
- 11: sp-rodent:\*
- 12: sp-virus:\*
- 13: sp-vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp-virus:\*
- 16: sp-bacteriopl:\*
- 17: sp-archaeop:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	51	14.6	340	11	Q8C608
2	34	9.7	159	11	Q9D146
3	10	2.9	366	16	Q8NMB9
4	10	2.9	366	16	Q8CSD9
5	9	2.6	415	16	Q8E582
6	8	2.3	150	10	Q8H386
7	8	2.3	242	5	Q8SV52
8	8	2.3	262	17	Q8S376
9	8	2.3	262	17	Q8U233
10	8	2.3	293	12	Q8O851
11	8	2.3	445	16	Q9KRL6
12	8	2.3	458	16	Q9KRL6
13	8	2.3	524	16	Q9Z4X2
14	8	2.3	1063	5	Q8SVT4
15	8	2.3	5922	5	Q8IIN2
16	7	2.0	69	10	Q9LRD3

17	7	2.0	73	11	Q55066	055066 apodemus fl
18	7	2.0	73	11	Q55065	055065 tokudata os
19	7	2.0	83	2	Q8KPU4	08KPU4 synechococ
20	7	2.0	95	6	Q97608	097608 didelphis m
21	7	2.0	95	6	Q97622	097622 wallabia bl
22	7	2.0	95	6	Q97607	097607 monodelphis
23	7	2.0	95	6	Q97667	097667 macropus eu
24	7	2.0	97	5	Q9XU93	09XU93 caenorhabd1
25	7	2.0	104	16	Q97J70	097J70 clostridium
26	7	2.0	107	16	Q916B0	0916B0 pseudomonas
27	7	2.0	107	16	Q8YX54	08YX54 anabaena sp
28	7	2.0	110	16	Q8P924	08P924 xanthomonas
29	7	2.0	112	4	Q9H1T4	09H1T4 homo sapien
30	7	2.0	115	11	Q99P43	099P43 mus musculu
31	7	2.0	116	11	Q8BRT7	08BRT7 mus musculu
32	7	2.0	117	6	Q18925	018925 trichosurus
33	7	2.0	117	10	Q9SXR0	09SXR0 oryza sativ
34	7	2.0	121	11	Q9JUR5	09JUR5 mus musculu
35	7	2.0	139	4	Q9H1T5	09H1T5 homo sapien
36	7	2.0	144	2	Q8RYT9	08RYT9 stigmatalia
37	7	2.0	149	6	Q8S022	08S022 capra hircu
38	7	2.0	163	17	Q9Y9S5	09Y9S5 aeropyrum p
39	7	2.0	164	5	Q9VNV7	09VNV7 dirosophila
40	7	2.0	164	6	Q77790	077790 bos taurus
41	7	2.0	164	13	Q90ZC2	090ZC2 calotes ver
42	7	2.0	175	6	Q95L21	095L21 ovis aries
43	7	2.0	177	17	Q95X99	095X99 sulfolobus
44	7	2.0	181	16	Q67372	067372 aquifex aeo
45	7	2.0	183	10	Q8L864	08L864 arabidopsis

## ALIGNMENTS

RESULT 1  
ID Q8C608 PRELIMINARY; PRT; 340 AA.

AC Q8C608:

DT 01-MAR-2003 (TREMBLrel. 23, Created)

DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)

DE Ngl interacting factor 3-like 1.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC SRRAIN-C57BL/6J; TISSUE-Testis;

RX MEDLINE=22354683; PubMed=12466851;

RA The FANTOM Consortium.

RA the RIKEN Genome Exploration Research Group Phase I & II Team;

RT "Analysis of the mouse transcriptome based on functional annotation of

RL 60,770 full-length cDNAs.";

RL Nature 420:563-573(2002).

DR EMBL AK076788; BAC36481.1;

SQ SEQUENCE 340 AA; 37822 MW; 82DC63DB90B68B4E CRC64;

QY Query Match 14.6%; Score 51; DB 11; Length 340;

Best Local Similarity 100.0%; Pred. No. 1.3e-44;

Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 MDKALLSLINDFASLSFAESMDNGLVPSPPHYNTLFLTNDLTFEVM 51

27 MDKALLSLINDFASLSFAESMDNGLVPSPPHYNTLFLTNDLTFEVM 77

RESULT 2  
ID Q9D146 PRELIMINARY; PRT; 159 AA.

DT 01-JUN-2001 (TREMBLrel. 17, Created)

DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)  
 DE 1110030G24RIK protein.  
 GN 1110030G24RIK.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Embryo;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aitawa K., Itawa M., Nishi K., Kiyosawa H., Kondo S., Yamano I.,  
 RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schiraldi L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Mashio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamlya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
 RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection."  
 RL Nature 409:685-690(2001).  
 DR EMBL: AK003978; BAB23102.1.  
 DR MGI: MGI:1913428; 1110030G24RIK.  
 DR InterPro: IPR002678; DUF34.  
 DR Pfam: PF01784; DUF34; 1.  
 SQ SEQUENCE 159 AA; 17559 MW; FAABCEB4FC165AC CRC64;  
  
 Query Match 9.7%; Score 34; DB 11; Length 159;  
 Best Local Similarity 100.0%; Pred. No. 4.7e-27;  
 Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 OY 203 QLYOKTEITLSLEKPLLTHTGMRGLCTLDESVSLA 236  
 DB 12 QLYOKTEITLSLEKPLLTHTGMRGLCTLDESVSLA 45  
  
 RESULT 3  
 OBNMB9 PRELIMINARY; PRT; 366 AA.  
 AC OBNMB9;  
 DT 01-OCT-2002 (TREMBLrel. 22, Created)  
 DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE Hypothetical protein MW1511.  
 GN MW1511.  
 OS Staphylococcus aureus (strain MW2).  
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
 OX NCBI\_TaxID=196620;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22040717; PubMed=12044378;  
 RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,  
 RA Nagai Y., Iwana N., Asano K., Naito T., Kuroda H., Cui L.,  
 RA Yamamoto K., Hiratsuka K.;  
 RT "Genome and virulence determinants of high virulence community-  
 acquired MRSA."  
 RL Lancet 359:1819-1827(2002).  
 DR EMBL: AP004827; BAB95376.1;  
 DR InterPro: IPR002678; DUF34.  
 DR Pfam: PF01784; DUF34; 1.  
 DR TIGRfams: TIGR00486; TIGR00486.1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 366 AA; 41153 MW; FALF28762D878C79 CRC64;

Query Match 2.9%; Score 10; DB 16; Length 366;  
 Best Local Similarity 100.0%; Pred. No. 0.17;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 OY 19 AESMDNVGLL 28  
 DB 19 AESMDNVGLL 28  
  
 RESULT 4  
 OBCSD9 PRELIMINARY; PRT; 366 AA.  
 AC OBCSD9;  
 DT 01-MAR-2003 (TREMBLrel. 23, Created)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE Conserved hypothetical protein.  
 GN SEI246.  
 OS Staphylococcus epidermidis  
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
 OX NCBI\_TaxID=1282;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 12228;  
 RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,  
 RA Chen Z., Wen Y.;  
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AE016748; AA004845.1;  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 366 AA; 41639 MW; BF73A88EE77462A1 CRC64;  
  
 Query Match 2.9%; Score 10; DB 16; Length 366;  
 Best Local Similarity 100.0%; Pred. No. 0.17;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 OY 19 AESMDNVGLL 28  
 DB 19 AESMDNVGLL 28  
  
 RESULT 5  
 OBE5B2 PRELIMINARY; PRT; 415 AA.  
 AC OBE5B2;  
 DT 01-MAR-2003 (TREMBLrel. 23, Created)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE Hypothetical protein.  
 GN GBS1120.  
 OS Streptococcus agalactiae (serotype III).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=216495;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NEM316 / Serotype III;  
 RX MEDLINE=22242508; PubMed=12354221;  
 RA Glaser P., Rusniok C., Buchrieser C., Chevalier F., Frangeul L.,  
 RA Msadek T., Zouine M., Couve E., Talloni L., Poyart C., Tieu-Cuot P.,  
 RA Kunst F.;  
 RT "Genome sequence of Streptococcus agalactiae, a pathogen causing  
 invasive neonatal disease."  
 RL Mol. Microbiol. 45:1499-1513(2002).  
 DR EMBL: AL766848; CAD46779.1;  
 DR Sagalistic: gbs1120;  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 415 AA; 47213 MW; 20029CB546FEA77F CRC64;

Query Match 2.6%; Score 9; DB 16; Length 415;  
 Best Local Similarity 100.0%; Pred. No. 2.1;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 281 GVEADLYLT 289  
 |||||  
 DB 176 GVEADLYLT 184

## RESULT 6

ID 08H3B6 PRELIMINARY; PRT; 150 AA.

AC 08H3B6:  
 DT 01-MAR-2003 (TREMBLrel. 23, Created)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
 DE 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 GN P0473C09.14 protein.  
 OS Oryza sativa (Japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 ON NCBI\_Taxid-39947;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Nipponbare;  
 RA Sasaki T., Matsumoto T., Katayose Y.;  
 RT "Oryza sativa nipponbare(GAS) genomic DNA, chromosome 7, PAC  
 clone:P0473C09.";  
 RT Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
 RL EMBL: AP005188; BAC20838.1; -; 2E53BEC83941A43F CRC64;  
 SQ SEQUENCE 150 AA; 16837 MW; 2E53BEC83941A43F CRC64;

Query Match 2.3%; Score 8; DB 10; Length 150;  
 Best Local Similarity 100.0%; Pred. No. 9.3;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 118 LGACTSRP 125  
 |||||  
 DB 58 LGACTSRP 65

## RESULT 7

ID 08SV52 PRELIMINARY; PRT; 242 AA.

AC 08SV52:  
 DT 01-JUN-2002 (TREMBLrel. 21, Created)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
 DE 01-OCT-2002 (TREMBLrel. 22, Last annotation update)  
 DE Hypothetical protein ECU07\_0100.  
 GN ECU07\_0100.  
 OS Encephalitozoon cuniculi.  
 OC Eukaryota; Fungi; Microsporidia; Unikaryoniidae; Encephalitozoon.  
 ON NCBI\_Taxid-6035;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GB-M1;  
 RA Genoscope;  
 RT Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GB-M1;  
 RA MEDLINE-21576510; PubMed-11719806;  
 RA Katsinka M.D., Duprat S., Cornillot E., Metenier G., Thomarac F.,  
 RA Prentiss G., Barbe V., Peyret P., Brothier P., Winkler P.,  
 RA Delbac F., El Alaoui H., Peyret P., Saurin W., Gouy M.,  
 RA Weissenbach J., Vivares C.P.;  
 RT "Genome sequence and gene compaction of the eukaryote parasite  
 Encephalitozoon cuniculi.";  
 RT Nature 414:450-453(2001).  
 DR EMBL: AL590447; CAD25542.1; -;  
 DR InterPro: IPR002678; DUF34.  
 DR Pfam: PF01784; DUF34; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 242 AA; 26877 MW; D95448A74B97B7D6 CRC64;

Query Match 2.3%; Score 8; DB 5; Length 242;

Best Local Similarity 100.0%; Pred. No. 15;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 289 TGESSHHD 296  
 |||||  
 DB 182 TGESSHHD 189

## RESULT 8

ID 058376 PRELIMINARY; PRT; 262 AA.

AC 058376:  
 DT 01-AUG-1998 (TREMBLrel. 07, Created)  
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
 DE 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE Hypothetical protein PH0642.  
 GN PH0642.  
 OS Pyrococcus horikoshii.  
 OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
 OC Pyrococcus.  
 ON NCBI\_Taxid-53953;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=OT3;  
 RX MEDLINE-98344137; PubMed-9679194;  
 RA Kawarabayashi Y., Sawada M., Horikawa H., Halkawa Y., Hino Y.,  
 RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,  
 RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,  
 RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,  
 RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,  
 RA Masuuchi Y., Shizuya H., Kikuchi H.;  
 RT "Complete sequence and gene organization of the genome of a hyper-  
 thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";  
 RT DNA Res. 5:55-76(1998).  
 DR EMBL: AP000003; BAA29733.1; -;  
 DR InterPro: IPR003010; Ntlase/CNhydase.  
 DR Pfam: PF00795; CN\_hydrolase; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 262 AA; 29806 MW; F416407E6822ADF2 CRC64;

Query Match 2.3%; Score 8; DB 17; Length 262;  
 Best Local Similarity 100.0%; Pred. No. 16;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 87 IRLAENRV 94  
 |||||  
 DB 183 IRLAENRV 190

## RESULT 9

ID 08U233 PRELIMINARY; PRT; 262 AA.

AC 08U233:  
 DT 01-JUN-2002 (TREMBLrel. 21, Created)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
 DE 01-OCT-2002 (TREMBLrel. 22, Last annotation update)  
 DE Hypothetical protein PF1017.  
 GN PF1017.  
 OS Pyrococcus furiosus.  
 OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
 OC Pyrococcus.  
 ON NCBI\_Taxid-2261;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VCI / DSM 3638 / ATCC 43587 / JCM 8422;  
 RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;  
 RT "The complete sequence of the Pyrococcus furiosus genome.";  
 RT Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AE010213; AAI81141.1; -;  
 DR InterPro: IPR003010; Ntlase/CNhydase.  
 DR Pfam: PF00795; CN\_hydrolase; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 262 AA; 29863 MW; F1B652E7F72F8F60 CRC64;

Query Match 2.3%; Score 8; DB 17; Length 262;  
 Best Local Similarity 100.0%; Pred. No. 16;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 87 IRLAENV 94  
 |||||  
 183 IRLAENV 190

## RESULT 10

ID 080851 PRELIMINARY; PRT; 293 AA.

AC 080851;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE 3a protein.  
 OS Hydrangaea mosaic virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Bromoviridae;  
 OC Ilarvirus.  
 OC NCBI\_TaxID=42812;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RX MEDLINE=96290233; PubMed=8725121;  
 RA Ge X., Scott S.W.;  
 RT "The nucleotide sequence of hydrangaea mosaic virus RNA 3 exhibits  
 RT similarity with the RNA 3 of tobacco streak virus.";  
 RL Virus Res. 40:57-63(1996).  
 DR EMBL: U35145; AAA80344.1;  
 DR InterPro: IPR002538; Bromo\_MP.  
 DR Pfam: PF01573; Bromo\_MP; 1.  
 SQ SEQUENCE 293 AA; 32112 MW; 734612E4A1934A75 CRC64;

Query Match 2.3%; Score 8; DB 12; Length 293;  
 Best Local Similarity 100.0%; Pred. No. 17;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 201 NKOLYORT 208  
 |||||  
 176 NKOLYORT 183

## RESULT 11

ID 09KRL6 PRELIMINARY; PRT; 445 AA.

AC 09KRL6;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE Agglutination protein.  
 GN VCI621.  
 OS Vibrio cholerae.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
 OC Vibrionaceae; Vibrrio.  
 OC NCBI\_TaxID=666;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RX STRAIN=El Tor N16961 / Serotype O1;  
 RX MEDLINE=20406833; PubMed=10952301;  
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,  
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,  
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
 RA Monodave M.D., Vamathevan J., Bass S., Qin H., Dragol I., Sellers P.,  
 RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,  
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
 RA Fraser C.M.;  
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio  
 RT cholerae.";  
 RL Nature 406:477-483(2000).  
 DR EMBL: AE004240; AAF94774.1;  
 DR TIGR: VCI621;  
 DR InterPro: IPR003423; OEP.

DR Pfam: PF02321; OEP; 2.  
 KW Complete proteome.  
 SQ SEQUENCE 445 AA; 49375 MW; A38FAF1E106E7150 CRC64;

Query Match 2.3%; Score 8; DB 16; Length 445;  
 Best Local Similarity 100.0%; Pred. No. 26;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 SKAPNPT 136  
 |||||  
 272 SKAPNPT 279

## RESULT 12

ID 08ECD1 PRELIMINARY; PRT; 458 AA.

AC 08ECD1;  
 DT 01-MAR-2003 (TREMBLrel. 23, Created)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE Flagellar biosynthetic protein FlhF.  
 GN FLH OR SO3212.  
 GN SHW OR SO3212.

OS Shewanella oneidensis.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;  
 OC Alteromonadaceae; Shewanella.  
 OC NCBI\_TaxID=70863;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RX STRAIN=MR-1;  
 RX MEDLINE=22297686; PubMed=12368813;  
 RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Galdo E.J., Nelson W.C.,  
 RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,  
 RA Meyer T., Tsaplin A., Scott J., Beanan M., Brinkac L., Daugherty S.,  
 RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,  
 RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,  
 RA Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,  
 RA Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A.,  
 RA Feldblyum T.V., Smith H.O., Venter J.C., Nelson K.H., Fraser C.M.;  
 RT "Genome sequence of the dissimilatory metal ion-reducing bacterium  
 RT Shewanella oneidensis.";  
 RL Nat. Biotechnol. 20:1118-1123(2002).  
 DR EMBL: AE015759; AAN56211.1;  
 DR TIGR: SO3212;  
 DR Complete proteome.  
 SQ SEQUENCE 458 AA; 49753 MW; 43FEE6E63435BF7A CRC64;

Query Match 2.3%; Score 8; DB 16; Length 458;  
 Best Local Similarity 100.0%; Pred. No. 26;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 229 LDESYSLA 236  
 |||||  
 387 LDESYSLA 394

## RESULT 13

ID 0924X2 PRELIMINARY; PRT; 524 AA.

AC 0924X2;  
 DT 01-MAY-1999 (TREMBLrel. 10, Created)  
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE Putative transcriptional regulator.  
 GN SCO3209 OR SCE8.02.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Streptomycinae; Streptomycetaceae; Streptomyces.  
 OC NCBI\_TaxID=1902;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RX STRAIN=A3(2);  
 RC Saunders D., Harris D.;  
 RA Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.

```

RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RN SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kleser H.M., Denapate D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome."
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RN SEQUENCE FROM N.A.
RC STRAIN-A3(2)/M145;
RX MEDLINE=21996410; PubMed=1200953;
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabbittowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wierozorek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
coelicolor A3(2)."
RL Nature 417:141-147(2002).
CC - SIMILARITY: BELONGS TO THE ICLR FAMILY OF TRANSCRIPTIONAL
REGULATORS.
CC EMBL: AL393115; CAB38580.1; -
DR InterPro: IPR001092; HTH_basic.
DR InterPro: IPR005471; HTH_ICLR.
DR Pfam: PF01614; ICLR; 2.
DR SMART: SM00346; HTH_ICLR; 2.
DR PROSITE: PS00038; HTH_1; 1.
KW DNA-binding; Transcription regulation; Complete proteome.
SQ
SEQUENCE 524 AA; 55353 MW; 2D96319552CEB307 CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 16; Length 524;
Pred. No. 30;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 229 LDESVSILA 236
DB 106 LDESVSILA 113

RESULT 14
OBSVIA
ID OBSVIA PRELIMINARY; PRT; 1063 AA.
AC OBSVIA;
DT 01-JUN-2002 (TREMblrel. 21, Created)
DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)
DT 01-OCT-2002 (TREMblrel. 22, Last annotation update)
DE Hypothetical protein EC005_1100.
GN EC005_1100.
OS Eucephalitozoon cuniculi.
OC Eukaryota; Fungi; Microsporidia; Unikaryonidae; Eucephalitozoon.
OX NCBI_TaxID=6035;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN-GB-M1;
RA Genoscope;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN-GB-M1;
RX MEDLINE=21576510; PubMed=11719806;
RA Kalinka M.D., Duprat S., Cornillot E., Metenier G., Thomarat F.,
RA Prenster G., Barbe V., Peyretallade E., Brotlier P., Wincker P.,

```

```

RA Delbac F., El Alaoui H., Peyret P., Saurin W., Gouy M.,
RA Weissbach J., Vivares C.P.;
RT "Genome sequence and gene compaction of the eukaryote parasite
RT Eucephalitozoon cuniculi."
RL Nature 414:450-453(2001).
DR EMBL: AL590445; CAD26630.1; -
DR InterPro: IPR001841; znf_ring.
DR SMART: SM00184; RING; 1.
DR PROSITE: PS50089; 2F_RING_2; 1.
KW Hypothetical protein.
SQ
SEQUENCE 1063 AA; 122205 MW; 2E8A8485D863585 CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 5; Length 1063;
Pred. No. 57;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 LKALLSSL 10
DB 360 LKALLSSL 367

RESULT 15
OBSVIA
ID OBSVIA PRELIMINARY; PRT; 5922 AA.
AC OBSVIA;
DT 01-MAR-2003 (TREMblrel. 23, Created)
DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE Hypothetical protein.
GN PF1_0528.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN-3D7;
RX MEDLINE=22255705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S.G., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Valdeya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum."
RL Nature 419:498-511(2002).
DR EMBL: AE014837; AAN35722.1; -
KW Hypothetical protein.
SQ
SEQUENCE 5922 AA; 710211 MW; 2066B6DBA444B225 CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 5; Length 5922;
Pred. No. 2,8e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 333 NKINIITS 340
DB 5795 NKINIITS 5802

Search completed: August 22, 2003, 15:19:18
Job time : 101 secs

```

**THIS PAGE BLANK (USPTO)**